AMENDMENTS TO THE CLAIMS

1. (currently amended) A computer-implemented method for improving information retrieval, classification, indexing, and summarization, comprising:

identifying a compound document as a coherent body of hyperlinked material on a single topic as created by a number of collaborating authors, wherein the body of hyperlinked material is distributed over a plurality of URLs;

analyzing the content and structure of the compound document to find a preferred entry point for the compound document;

processing the compound document as a whole, including at least one of indexing, classification, and retrieval; and

processing the compound document from the entry point, including at least one of creating at least one of presentation of results from retrieval, summarization, and classification,

wherein the identifying includes observing results of running a first number of heuristics on the body of hyperlinked material and related hyperlinks, wherein running the first number of heuristics includes

identifying hyperlinks that link within a same directory,

identifying hyperlinks that contain linguistic structures that indicate relationships between document parts,

identifying external hyperlinks to same places,

identifying at least one of: similar creation dates and similar last-modified dates,

identifying individual URLs having similar structure indicating an order of inclusion in the compound document,

identifying a link structure of "wheel" form,

wherein the analyzing includes observing the results of running a second number of heuristics on component documents in the compound document and related hyperlinks, wherein running the second number of heuristics includes

identifying specific filenames that define entry points, including at least one of: "index" and "default",

identifying a particular component document in the compound document as a suitable entry point because the component document has several in-links, wherein the in-links are from outside the compound document,

determining a measure of vector distances along intra-document links between a particular component document and all other component documents in the compound document,

determining whether a URL has links pointing to longer URLs having common directory components followed by different ending directory components, wherein the ending directory components contain specific identifying information, and

wherein the analyzing includes combining the results of running the second number of heuristics on the various component documents in the compound document, wherein the results include numerical scores and the combining includes a weighted averaging of the numerical scores into an overall score, and the maximum overall score determines the preferred entry point.

2-21. (cancelled)